Canadian Journal of Philosophy (2020), 50: 3, 395–412
doi:10.1017/can.2019.44

ARTICLE

Kant on Cognizing Oneself as a Spontaneous Cognizer

Markus Kohl

Department of Philosophy, University of North Carolina, Chapel Hill
Email: mkohl17@email.unc.edu

Abstract
I examine a range of issues concerning Kant's conception of cognitive spontaneity. I consider whether we can cognize or know ourselves as spontaneous cognizers, and why Kant seems to regard the notion of cognitive spontaneity as less problematic than the idea of moral spontaneity. As an organizing theme of my discussion, I use an apparent tension between the A-edition and the B-edition of the first *Critique.* Against common interpretations, I argue that in the B-edition Kant does not revoke his claim that we can cognize, and (perhaps) even know, that our noumenal selves are absolutely spontaneous cognitive agents.

Keywords: Kant; spontaneity; freedom of thought; freedom of will; apperception; self-knowledge

Introduction

Kant constantly invokes spontaneity as a central feature of both our theoretical cognition and our moral agency. He (apparently) takes our cognitive spontaneity for granted, whereas he acknowledges that the idea of moral spontaneity raises significant difficulties. What explains this difference? Are cognitive and moral spontaneity two species of one kind or are they completely dissimilar? Does our cognitive spontaneity belong to our empirical self or (like our moral spontaneity) to our noumenal self? If the latter, how could we cognize or know that we are spontaneous cognizers? In this essay, I aim to shed new light on these questions.

As an organizing theme of my discussion, I shall use an apparent tension between the 1781 A-edition and the 1787 B-edition of the *Critique of Pure Reason*[1]. In the A-edition Transcendental Dialectic (A546–47), Kant allows that a human being can theoretically "cognize itself through pure apperception" (the "I think") "as a merely intelligible object" whose cognitive capacities (understanding and reason) must be "distinguished … from all empirically conditioned powers." These spontaneous capacities enable me to spontaneously determine myself in "actions and inner determinations." But in the B-Paralogisms, Kant argues that it is "impossible" for the thinking self "to cognize it[s] [existence] as noumenon" (B430): the apperceptive self-awareness of my "sheer spontaneity" gives me no noumenal self-cognition because the application of this spontaneity "to sensible intuition … would … be demanded if I wanted to cognize myself," and sensible "intuition always makes available the object … merely as appearance" (B429). Kant now stresses that we require our *practical* awareness of the moral law as an "occasion for presupposing ourselves to be … self-determining in [our] existence," since only the moral law reveals "a spontaneity through which our actuality is determinable without the need of conditions of empirical intuition." (B431–32).

[1]Quotations from Kant's works—apart from the *Critique of Pure Reason* which is cited according to the standard A/B pagination—cite the volume and page number of the Academy edition. Translations are based on the Cambridge Edition of the works of Kant.

The Author(s) 2020. Published by Canadian Journal of Philosophy.

A common response to this seeming conflict is that by 1787 Kant had given up the belief that we can theoretically cognize our noumenal capacity for spontaneous self-determination.[2]

In this essay, I aim to defuse the conflict between these passages and thereby illuminate Kant's account of spontaneous self-determination. I will proceed as follows. In section 1, I argue that Kant's varying claims about whether we may cognize our cognitive spontaneity can be reconciled by distinguishing between a wide and a narrow notion of cognition. In sections 2 and 3, I argue that for Kant our cognitive and moral spontaneity involve two different kinds of self-determination; only the idea of moral spontaneity allows for truly *causal* self-determination, and therefore this idea is harder to justify than the noncausal notion of cognitive spontaneity. Despite this difference, cognitive spontaneity is a species of absolute, causally unconditioned spontaneity (transcendental freedom). In section 4, I argue that in Kant's critical view, we have a theoretical basis for affirming, and (perhaps) even knowing, that our noumenal selves are absolutely spontaneous cognitive agents.

1.

I first address the seeming conflict between Kant's statements that we can and cannot cognize our noumenal thinking self. Here I want to consider the fact that Kant has two (related) notions of cognition.[3] Cognitions in a narrow sense (N) are representational states that combine concepts with sensible intuitions and thereby, unlike "empty" concepts or "blind" intuitions, inform us about the properties of actual ("given") objects (A50–52/B74–76; B146–47). Kant also has a wider notion of cognition (W) that he contrasts with sensations. Empty concepts and blind intuitions qualify as cognitions(W) because they are conscious representations that relate to objects. Empty concepts relate to objects in the sense that they have intentional "aboutness": they represent something distinct from the occurrent mental state of conceiving. This intentional relation to an object need not involve any existential commitment, though via cognition(W) we can at least represent something *as* existing apart from (external to) our mental state. Thus, all purely intellectual representations, including even the transcendent "general cognitions of reason" (A421/B448), are cognitions(W). By contrast, sensations like feelings of pleasure and pain have no cognitive content; as mere occurrent modifications of some particular mind, they are not intentionally related to (are not representations *of*) any object or property (A320/B376; *cf.* vol. 9, 91; vol. 28, 471).

In the B-Paralogisms, Kant clearly uses "cognition(N)" when he argues that since pure apperception lacks intuitive content it cannot yield cognition of our noumenal self. This does not contradict his claim (at A546/B574) that through pure apperception a human being can cognize itself as an intelligible object if we assume that here Kant uses "cognition(W)." This assumption does not seem excessively charitable: since Kant repeatedly denies that pure apperception has the intuitive content that would be required for self-cognition(N) (A382; A400; B159), we can conjecture that his appeal to self-cognition through pure apperception involves the broader notion of cognition(W) that does not require intuition. In saying that pure apperception is a kind of cognition despite its lack of intuitive content, Kant stresses that pure apperception is not a private, nonintentional mental state but a universally shared, intellectual representation through which every thinking being conceives itself as the transcendental subject of thoughts (A346/B404; B423).

Since cognition(W) is defined as a conscious representation that relates to an object, pure apperception can only yield cognition(W) of our noumenal self if it represents our thinking self as a noumenal object. Kant's conception of object-representation allows for this possibility. To be sure, he has a narrow conception of object-representation that is only satisfied by mental states which combine intuitive and conceptual components into determinate, informative cognition(N) of

[2]See Ameriks (2000, 217; 2003, 166–67). See also Hogan (2009, 379) and Keller (1998, 159).

[3]For this point, I am indebted to Grüne (2009, 29–30) and the discussion in Watkins and Willaschek (2017).

phenomena and their spatiotemporal properties.[4] When Kant denies that pure self-consciousness represents the thinking self as an object (e.g., at B407), he uses this narrow notion of an object that designates a determinate object of cognition(N). But if one abstracts from all sensible data for cognition(N), one can still represent an indeterminate "something" via the pure concept of "an object in general" that is the highest concept in all cognition (A290/B346; vol. 28, 544). Accordingly, pure apperception indeterminately represents the thinking self as a "something in general" (A400; *cf.* A345–46/B403–4; A383; B421). Because pure apperception intentionally relates to a nonsensible ("merely intelligible"; A546/B574) something that is conceived as distinct from my occurrent mental states, it yields cognition(W) of myself as a nonspatiotemporal, noumenal object. (I further clarify and defend this point below, when I consider Michelle Grier's account.)

A problem for the reading that I am proposing arises from Kant's B-edition assertion that through the "I think," "I am conscious of myself not as I appear to myself, nor as I am in myself, but only that I am" (B157; *cf.* B429). Here, one might argue, Kant denies that pure apperception yields cognition of an intelligible object (of myself "as I am in myself") *even if* "cognition" and "object" are understood in the wide sense (W).[5]

However, here we must consider how Kant argues for his claim that in pure apperception I am not conscious of myself "as" I am in myself. He first stresses that I require sensible intuition to cognize(N) how I exist as a thinking being, and from that he infers that pure, nonsensible apperception does not yield "cognition of oneself." "I therefore have no cognition of myself as I am, but only as I appear to myself" (B158). Thus, when Kant denies that pure apperception gives me a consciousness of myself as I am in myself, he only seeks to stress that pure apperception does not give me the kind of self-cognition(N) that requires sensible intuition. In pure apperception I am not conscious of myself "as" I am in myself because my purely intellectual self-cognition(W) provides no determinate information about the manner in which I exist as a thing in itself (B420). In the B-edition, Kant still accepts that pure apperception represents the thinking self as an "object in general" (B158, *cf.* B429) and that it provides an indeterminate self-characterization: "I exist as an intelligence that is merely conscious of its faculty for combination." Although I cannot cognize(N) my spontaneous intellect, "yet this spontaneity is the reason I call myself an intelligence" (B158); "the I that I think … differ[s] from the I that intuits itself" since the I that I think is the "I as intelligence and thinking subject" (B155). Our self-awareness as an *intelligence* necessarily relates to our noumenal self: to consider human beings as noumena *is* to consider them according to the capacities they have as an intelligence (vol. 6, 226).[6] Thus, Kant's remarks about pure apperception in the B-edition seem compatible with his claim in the A-edition Dialectic: my pure self-consciousness gives me indeterminate cognition(W) of myself as a spontaneous, noumenal intelligence.

However, some commentators argue that in Kant's critical doctrine pure apperception cannot represent a noumenal object. Here I want to focus on Michelle Grier's influential reading. She makes two claims that are especially relevant for our topic. First, the "I" that is represented in pure apperception relates merely to the formal or logical subject of thought, which "is not an object of any sort" (2001, 168–69). Second, "the conflation of the 'I' of apperception with a fictitious object" (170). occurs under the influence of a transcendental illusion, when pure reason hypostatizes the logical subject of thought into an "illusory idea of the soul" as a thing in itself or noumenon (168–69).

---

[4]To clarify: the very concept of determinate object-representation via cognition(N) need not entail that there is some *sensible* intuition that relates to some *phenomenal* object. Perhaps intellectual intuition of noumena would also qualify as cognition(N). But cognition(N) involving sensible data is the only cognition(N) that our human (or any finite) mind may obtain.

[5]See Kitcher (1984, 121).

[6]For the analytic link between "intelligence" and "noumenal being," see vol. 4, 452, 457–59; vol. 28, 583, 773, 1115–16; vol. 29, 1020; see also Puls (2015, 184–85). Against Schönecker (1999, 287–89), Puls argues that "intelligence" designates not the capacity for theoretical judgment but (primarily) the capacity for moral agency. However, passages such as B158; vol. 6, 226; and vol. 4, 452 suggest that the term designates both capacities.

When Grier claims that the pure concept of "I" represents no object, she means that it does not represent a determinate object of intuitive cognition(N) because "it is unacceptable to think of this 'I' as an object given in … intuition" (161). She accepts that the transcendental subject of thought can be thought as an "object in general" (160, 170). Accordingly, in the passages she cites (B407, B409–10), Kant makes the same point as in the B-deduction (B157–58): I cannot cognize(N) myself as a *determinate* object (that has specific properties or characteristics) without intuition. But for Grier the purely conceptual representation of a nonsensible object in general is the thought of a mere pseudo-object that lacks ontological import. This informs her claim that an illusion of reason is needed to create the fictitious impression that pure apperception represents the thinking self as a thing in itself. These claims are based on Grier's view that Kant's appeal to things in themselves involves no commitment to the existence of nonsensible objects: "… critically understood, 'things in themselves' are not objects at all but certain ways of considering objects" in abstraction from the sensible conditions of human cognition(N).[7] For Grier, "the consideration of things as they are in themselves is basically an illusory way of representing appearances" (2001, 278).

My interpretation presupposes that this deflationary reading of Kant's idealism is mistaken, namely, that Kant posits nonsensible things in themselves in addition to sensible appearances.[8] In the context of this essay, I cannot conclusively support this presupposition. I can only note that Kant often makes claims that tell against a deflationary reading.[9] He stresses that "… beings of the understanding certainly correspond to sensible beings …" (B308). If one accepts the existence of sensible appearances, one must accept not merely (as proponents of deflationary readings suggest) the *concept* of nonsensible beings but "the existence of things in themselves" that "underlie the appearances" (vol. 4, 315) as their "true correlate" (A30/B45) (see also Bxx; A538/B566; A496/B524; vol. 4, 355, 361).

The link between a metaphysical reading of Kant's idealism and my reading of his doctrine of self-cognition is as follows.[10] On a metaphysical reading, Kant holds that for everything that we determinately cognize(N) as a sensible appearance, there exists some nonsensible being that we can represent only indeterminately via purely conceptual cognition(W), "as a something in general outside of our sensibility" (B307). Here Kant uses his abovementioned indeterminate object-concept in conjunction with his negative concept of noumena to represent something in general that does *not* conform to our forms of sensibility. This is also a negative or indeterminate conception of an intelligible being since the "intelligible [is] that in an object of sense which is not itself appearance" (A538/B566). When we (negatively) represent intelligible beings, we refer to those (unknown, indeterminate) features of things that differ from the sensible constitution that they have as appearances. As I read Kant's views on self-cognition, we can determinately cognize(N) ourselves as sensible appearances and we also have purely conceptual cognition(W) of our thinking self as an object in general whose constitution does not conform to our sensibility and which is

[7]Grier (2001, 90). Here she defers to Allison's interpretation. Allison also reads Kant as warning against conflating "the indeterminate concept of a thinking being" with a "transcendent entity" (2004, 350). Similarly, for Rosefeldt (2000, 73–75) the pure "I" refers to a nonsensible being only qua logical pseudo-object, which he understands as an *ens rationis*. This point seems based on his general view that *all* things considered apart from our sensibility (noumena in the negative sense) are mere *entia rationis* since we cannot cognize whether or not any nonsensible objects exist.

[8]Here my reading is similar to that of Martin (1969, 207–8), who argues (against Cohen and Natorp) that a deflationary interpretation of the transcendental subject of thought is based on an implausibly deflationary understanding of Kant's appeal to things in themselves.

[9]For detailed criticisms of deflationary readings, see Allais (2015, 77–97), Ameriks (2003, 73–9, 103–5), and my own discussion concerning Kant's views on freedom in Kohl (2016).

[10]The label "metaphysical reading" covers a variety of interpretations. For my purposes, it is sufficient to note that a metaphysical reading is not committed to the idea that noumena and phenomena are separate entities. See Allais (2015, 3–36), Ameriks (2003, 35–38), and Kohl (2016). Marshall (2010; 2013) forcefully argues this point with regard to the human self, though I am not sure I follow his extension of his arguments beyond the case of the self.

therefore an intelligible object.[11] For instance, in the abovementioned Dialectic passage (A546–47/B574–75), Kant distinguishes between the sensible cognition(N) that we have of ourselves as empirical phenomena and the purely intellectual apperceptive cognition(W) that we have of ourselves as intelligible objects.[12]

Now, Grier is right to stress that for Kant the I of pure apperception designates a merely logical subject. But in Kant "logical" characterizations may well refer to real actions or things.[13] Pure apperception (unlike inner sense) involves an action awareness, "a consciousness of what the human being does" (vol. 7, 161), an "intellectual representation of the self-activity of a thinking subject" (B278). Moreover, Kant specifies "this logical meaning of the I" through the notion of inherence: the "I" refers to myself as the transcendental bearer of all my thoughts (A349–50). The pure "I" is a *merely* logical notion because it supplies "no determinate predicate" for conceiving this bearer (A479/B507), i.e., no predicate like permanence, persistence through change, *ultimate* subjecthood (where what is conceived as a subject of inherence cannot, in turn, inhere in some more fundamental subject), or separability from material beings. Since the logical meaning of "I" abstracts from all such determinate predicates, Kant concludes that "apart from this logical meaning of I, we have no acquaintance with the subject in itself that grounds this I as a substratum, just as it grounds all thoughts" (A351). This implies that the logical meaning of "I" *does* relate to the thinking subject as a thing in itself, albeit without giving us any determinate information about how it is constituted in itself: "the I, as the general correlate of apperception … designates … a thing of undetermined meaning" (vol. 4, 542). Through the "I think," the thinking subject is "only transcendentally signified" as a "something in general"; this transcendental signification does not reveal "the least property" (A355) but nevertheless refers to a noumenon (*cf.* A358), namely, to "the real self, as it exists in itself, or the transcendental subject" (A492/B520). In the B-edition Paralogisms, Kant likewise says that "in the consciousness of myself in mere thinking I am the being itself, about which, however, nothing yet is thereby given to me for thinking" (B429); through the concept "I," I can "signify" myself as an object in itself but I cannot cognize the manner in which I exist as a thing in itself (B430).

The indeterminate apperceptive representation of myself as a nonsensible *something* does not entail the problematic characterizations of the thinking self that rational psychology aspires to because the concept of a "thing," as the highest concept in all cognition, is less determinate than even the concept of substance (B6; vol. 9, 97) that plays a crucial role in rational psychology.[14] While the conception of my thinking self as a *spontaneous intelligence* seems to go beyond the mere thought of

[11]One might argue that pure apperception cannot represent the self as a nonsensible object because Kant characterizes the "I think" as an *empirical* proposition (B423). However, it is unclear what he means by this. He might seek to stress the *contingency* of the existential assertion that is implied by the "I think," or (perhaps: *and*) he might intend to stress that apperceptive self-consciousness requires empirical intuitions as the data for spontaneous thinking activity (*cf.* B423; Allison [2004, 354]). These points are compatible with the idea that the "I think" indeterminately represents my thinking self as an intelligible being. Longuenesse (2017, 87–88) distinguishes between (1) the "I think" qua "pure action-awareness" that (indeterminately) represents the thinking being in itself and (2) the indeterminate empirical intuition that I think. I am only concerned with (1). A referee for this journal suggests that perhaps Kant here only means that the "I think" *seems like* an empirical proposition if one adopts Lockean presuppositions (which Kant rejects). My reading is also compatible with (indeed, congenial to) this suggestion.

[12]I cannot consider Kant's intricate views on self-cognition in any further detail here due to limitations of space. In particular, I cannot discuss Kant's highly complex views on the positive content of empirical self-cognition(N). For recent discussion, see Chignell (2017).

[13]Here I concur with Kitcher's criticism of "logical" readings of apperception (1990, 94–95). For the point that Kant's "pure logic" concerns our actual mental faculties and operations, see also Longuenesse (1998, 74–78).

[14]It is controversial whether Kant can avoid the idea that the noumenal self is an (ultimate) substance. Ameriks argues that Kant must concede the substantiality of the noumenal soul (2000, 72–73; compare Wuerth [2014, 167–68]), but this forces Ameriks to dismiss Kant's first Paralogism argument as "spurious" (2000, 66–67). Longuenesse (2017) proposes that in Kant's view we are subjectively constrained to regard ourselves as ultimate substances, but we lack sufficient justification for judging that this self-conception is objectively correct.

something in general, Kant denies that pure consciousness of myself as a spontaneous intelligence gives me a determinate concept of how I exist as an intelligible being (B158). The ascription of spontaneity is not a target of Kant's criticism in the Paralogisms.[15] Hence it does not entail the substantive metaphysical characterizations of the thinking self that *are* the targets of Kant's criticism.[16] I will return to this issue in the final section.

I have argued that Kant's distinction between two senses of "object-cognition" allows us to see how his appeal to our apperceptive self-cognition as a spontaneous noumenon in the Dialectic is compatible with his seemingly more skeptical remarks in the B-edition Paralogisms. However, there is a further apparent tension between these passages, which I address in what follows.

2.

In the Dialectic, Kant asserts that pure apperception reveals our capacity for spontaneous "actions" and "inner determinations" that derive from our empirically unconditioned intellectual powers. In student notes on Kant's lectures, we find similar comments: "Now the 'I' proves that I myself act … I am conscious in myself of determinations and actions, and a subject who is aware of such actions and determinations has *libertatem absolutam*" (vol. 28, 266–69); "The intellectual capacity of cognition rests upon spontaneity or the capacity to determine oneself" (vol. 29, 881). In the 1783 *Prolegomena*, Kant stresses that "we are … completely free as to how we want to judge things" through "the understanding" (vol. 4, 290). But in the 1787 B-Paralogisms, Kant deems it "impossible" for the thinking self "to determine its kind of existence" via the "sheer spontaneity" of understanding—to regard ourselves as "self-determining in [our] existence" we must presuppose laws of *practical* reason (B430). Here Kant seems to stress that our pure theoretical self-awareness as thinking beings cannot reveal our capacity for free, spontaneous self-determination.[17] According to some commentators, this is his considered view: as Henry Allison puts it, Kant recognized that the "sheer spontaneity" of thought would make us rational *beings* but not *agents* because without the spontaneity of practical reason "our sense of agency would be illusory."[18]

It must be admitted that Kant's various remarks about our capacity for spontaneous self-determination are confusing.[19] Consider the following important but difficult passage from the B-deduction (B157–58):

> The I think expresses the act of determining my existence. (…) Now I do not have yet another self-intuition, which would give the determining in me … even before the act of determination, in the same way as time gives that which is to be determined, thus I cannot determine my existence as that of a self-active being, rather I merely represent the spontaneity of my thought, i.e., of the determining, and my existence always remains only sensibly determinable, i.e., determinable as the existence of an appearance.

[15]See Ameriks (2000, 190) and Rosefeldt (2000, 150–57).

[16]Grier (and Allison) might respond that at B426–427 Kant says that the conflation of the indeterminate concept of a thinking being with the idea of a pure intelligence betrays a transcendental illusion. However, if Kant meant that our self-awareness as a thinking intelligence is illusory, then *he* would have fallen prey to that illusion in the B-deduction (B155–58). The content of the illusory idea Kant targets at B426–27 is not merely that of a spontaneous intelligence, but the more determinate, controversial notion of a *separable* "pure intelligence" *that could exist apart from my empirical self* and consequently from all material, corruptible beings (see A741–42/B769–70).

[17]This is Ameriks's view (2000, 217; 2003, 166–67).

[18]Allison (1990, 63). For similar views, see Kitcher (1984; 1990; 2011), and Sellars (1970). In his later works (e.g., Allison [1996]), Allison seems open to the view that the spontaneity of thought is absolute. For this view, see also Pippin (1987) and my discussion in Kohl (2015). However, neither Pippin's essay nor my previous article consider how the absolute spontaneity reading can be reconciled with the B-edition Paralogisms.

[19]For Kitcher (1984, 121), these remarks are "individually perplexing and mutually inconsistent."

Here Kant makes (at least) three assertions:

(1) Through pure apperception, I represent myself as a self-active thinker whose spontaneity is "the determining in me."
(2) Through pure apperception I cannot determine my existence as a "determining" thinker.
(3) If I combine pure apperception with empirical intuition, I *can* determine my existence (or actuality) as an appearance.

Furthermore, as we saw, in the B-Paralogisms (B430) Kant claims:

(4) *Only* our awareness of practical laws "discloses a spontaneity through which our actuality is determinable without the need of conditions of empirical intuitions"; in this awareness "something is contained a priori that can serve to determine our existence."

Finally, we have the remarks from the lecture notes, the Dialectic and the *Prolegomena*:

(5) The capacity for theoretical cognition rests upon the spontaneous "capacity to determine oneself" via "actions and inner determinations" that result from the free exercise of our empirically unconditioned intellectual powers (theoretical reason and understanding).

I propose that a first step towards an interpretation on which (1) through (5) yield a coherent view is the disambiguation between two different senses of "spontaneous determination."

In the Analytic (A92–93/B124–26), Kant notes that there are two different senses "in which the [a priori] representation alone makes the object possible." In the first case, "the representation by itself" through "its causality by means of the will … produce[s] its object as far as its existence is concerned." I call this *practical-causal determination* (PCD). In the second case, a representation "does not produce its object as far as its existence is concerned" but "is still determinant of the object a priori if it is possible through it alone to cognize something as an object." Kant here means object-cognition(N) through conceptual synthesis of sensible intuition: "With us understanding and sensibility can determine an object only in combination." (A258/B314). I call this *theoretical-cognitive determination* (TCD).[20]

The distinction between PCD and TCD is illuminating with regard to (1) through (5). Part of what Kant means in (1) and in (5) is that through pure apperception I represent myself as a thinking being that has the capacity for spontaneous TCD of myself. In (2), he denies that pure apperception on its own enables the TCD of myself as a spontaneous thinker: since TCD of *any* object requires empirical intuition, the purely conceptual representation of my spontaneous thinking self is insufficient "to determine its kind of existence, i.e., to cognize it as a noumenon" (B430). In (3), Kant says that if I combine pure apperception with empirical intuition, then I can "determine my existence" in the sense of TCD: I can *cognitively* determine myself as a phenomenal object. In (4), Kant stresses that as moral agents, we may ascribe to ourselves a capacity for purely a priori self-determination in the sense of PCD. I can *causally* determine myself as a phenomenal object by making a choice which is potentially independent of all empirical desires: namely, when I exercise the capacity of pure reason to be "practical," to "determine the will by itself, independently of everything empirical" (vol. 5, 42). This is a form of causal *self*-determination because it determines my spatiotemporal actions and, thereby, the course of my sensible existence (*cf.* B430).

This analysis might remove the tension between (4) and (5). The claim (in [4]) that *only* our practical self-awareness under the moral law reveals our capacity to spontaneously determine our existence does not contradict the claim (in [1], [3], and [5]) that pure apperception *also* reveals a

[20]See Watkins and Willaschek (2017) for discussion of cognition(N) as the conceptual "determination" of intuition.

capacity to spontaneously determine our existence because these two claims concern two different types of spontaneous self-determination, namely, PCD and TCD. Moreover, since Kant never denies that our capacity for spontaneous self-determination in the theoretical-cognitive sense (whose exercise yields self-cognition[N]) requires empirical self-intuition, his claim that pure apperception reveals such a capacity does not conflict with his point in the B-edition that only the moral law reveals our capacity for spontaneous self-determination in a different, practical-causal sense that does *not* depend on empirical conditions.

If there is no genuine conflict between the two editions, then why does Kant so strongly emphasize in the rewritten B-Paralogisms, but not in the earlier A-Paralogisms, that we have a capacity for spontaneous self-determination that is independent of empirical conditions? I concede that there *is* a significant difference between Kant's 1781 and 1787 doctrine, but it concerns Kant's conception of the spontaneity of the will that is involved in PCD, not his conception of the spontaneity of thought that is involved in TCD. In the A-edition, Kant still holds a heteronomous view of moral motivation that he came to reject only by 1785 (in the *Groundwork*). According to that view, we are *not* capable of PCD independently of empirical conditions because our incentive to follow the moral law concerns our expectation of divine rewards in the afterlife. This motive depends on our empirical desire for happiness.[21] Hence, Kant's 1787 conception of PCD was not yet available to him in 1781. Once he has developed his doctrine of moral autonomy, he is in a position to stress that our capacity for PCD, unlike our capacity for TCD, does not depend on the limiting conditions of empirical intuition.[22] This is significant in Kant's system because it implies that practical reason has a decisive advantage over theoretical reason (*cf.* Bxxi).

There is a further potential conflict between (4) and (5): in (4), Kant stresses that the spontaneity of understanding depends on conditions of empirical intuition, whereas in (5), Kant refers to our empirically *unconditioned* theoretical faculties. I suggest that in (5) he means to stress the pure origin of intellectual representations that govern acts of TCD. Our theoretical categories and ideas are not abstracted from passively received empirical data; rather, they arise spontaneously from our purely intellectual self-activity (vol. 4, 452). The determinate *use or application* of pure concepts in acts of TCD (including acts of self-cognition[N]) always requires empirical data: "the empirical is only the condition of the application or the use of the purely intellectual capacity" (B423). The cognitive determination of objects that results from the application of purely intellectual representations to empirical data differs depending on whether these representations originate in pure reason or the understanding. Transcendent ideas of reason cannot cognitively determine objects or produce cognitions(N) directly, but they can contribute to acts of TCD by unifying the various cognitions(N) of the understanding into cognitive determinations of more complex objects or relations among objects. This is what Kant means when he says that reason "determines the understanding" (A547/B575) through its ideas. "For even if no object can be determined through [these ideas], they can still … serve the understanding as a canon for its extended and self-consistent use" (A329/B387. These ideas "determine the use of the understanding according to principles in the whole of an entire experience" (A321/B378).

I have argued that the distinction between TCD and PCD as two different types of spontaneous self-determination allows us to make coherent sense of Kant's confusing talk about spontaneous self-determination in (1) through (5). This distinction promises further interpretive benefits. If the "determining in me" that I represent via pure apperception is a capacity for cognitive *rather than* causal self-determination, then ascribing the capacity for TCD does not have the controversial metaphysical implications that arise with regard to PCD. With PCD, one ascribes a productive

[21]See A811/B839; A813/B841. See also A15; vol. 28, 1138, 1153.

[22]My reading can accept that when Kant developed his view of moral autonomy, he changed his strategy for proving that we have moral spontaneity of will. Thus, my reading can agree with Ameriks (2000; 2003) that there is a "great reversal" in Kant's moral metaphysics.

causal power that operates independently of all sensible conditions, which raises intricate questions about how we can extend the categories beyond the limits of our senses and how (or whether) we can demonstrate the (real) possibility of such a cause. The assumption that positing the capacity for spontaneous TCD lacks these controversial metaphysical implications might explain why Kant's critical epistemology constantly appeals to our intelligible capacity for spontaneous TCD but then claims that we cannot theoretically prove even the (real) possibility of the causal spontaneity that we conceive through the idea of PCD (A557–58/B585–86). Furthermore, that assumption might also explain why Kant does not treat the "sheer" spontaneity of mind in the Paralogisms alongside metaphysically weighty features of the thinking self. However, before I can confidently endorse these welcome interpretive results, I must address a significant complication.

3.

According to my reading in section 2, when Kant says that the spontaneity of thought in conjunction with empirical intuition allows me to "determine" my actual existence he means that I can exercise my capacity for TCD to obtain self-cognition(N). This is the right interpretation for some central passages such as B157–58. However, when Kant designates the understanding as a spontaneous "capacity to determine oneself" (vol. 29, 881), his notion of spontaneous self-determination cannot be the same as *self*-cognition: a cognizing being spontaneously determines itself in *any* acts of TCD whether it cognizes(N) itself or some other object (e.g., a table). In any act of TCD, one exercises control over one's mental states by forming certain representations, combining them in a judgment, and adopting some doxastic attitude (e.g., an opinion that a table is round). Following Kant's language in (5), we might call this *inner determination*. The idea that our thinking activity in TCD spontaneously determines our inner, representational states seems like a causal notion: judgment is the "effect" of "actions" of the understanding (A295/B350); experience is the "product" of the workings of the understanding (A1); the synthetic unity of intuition which yields object-cognition(N) is "effected" by the understanding (A105).

But if we must regard TCD as a kind of *causal* self-determination after all, this threatens the interpretive benefits that I tried to reap from the disambiguation between the two types of spontaneous self-determination. A causal conception of TCD not only seems to undermine my suggestion that ascribing a capacity for spontaneous TCD involves less of a metaphysical commitment than ascribing a capacity for spontaneous PCD. It also threatens the idea that there is any decisive difference at all between TCD and PCD. To be sure, we can still say that the causality of the cognizing mind, unlike the causality of the will, depends on empirical conditions. But the significance of this contrast is unclear if the contrast concerns a difference between two nonsensible (hence nonnatural) forms of causal self-determination rather than a difference between a causal and a *merely* cognitive sense of spontaneous self-determination. Finally, the causal status that Kant apparently ascribes to TCD in (5) (when he designates our theoretical faculties as nonempirical capacities for producing actual cognitive states as our "inner determinations") may also refuel the tension between (4) and (5): for one might read Kant's remarks in the B-Paralogisms (in [4]) as emphasizing that *only* our moral will exhibits a nonempirical form of spontaneity through which "our actuality is determinable" in a causal sense.

Some commentators are eager to dismiss the claims that Kant makes in (5). In their view, TCD is indeed a causal capacity for determining our actual mental states, but this requires no metaphysically loaded claims about noumenal causality because the capacity for TCD belongs to the empirical causality of nature. Thus, Patricia Kitcher argues that "the I that thinks is the phenomenal self" (1990, 140) whose synthesizing cognitive activity is determined by natural laws,[23] laws which

[23] Accordingly, Kitcher argues (1984, 122) that our faculty of cognition lacks genuine spontaneity. Her recent work still denies that the thinking self is free or genuinely spontaneous (2011, 170).

"govern synthesis only as the law of gravity governs the movements of the planets" (1990, 83).[24] On this reading, the thinking activity which produces cognition(N) is part of the empirical order of nature and is thus itself a potential object of cognition(N).

However, Kant denies these claims, as we can see in (1) through (3) (B157–58). In (1), Kant says that through pure apperception I represent my spontaneity as the agent of TCD, "the determining in me": "the, I think, expresses the Actus of [cognitively] determining my existence" when I obtain self-cognition(N). Here Kant refers to the synthesizing activity that produces cognition(N). Kitcher construes this synthesis as a phenomenal process which, as such, can itself become an object of cognition(N). But in (2), Kant explicitly denies that the cognitively determining aspect of my self can become an object of cognition(N): I am "merely conscious" of the spontaneity that is the "determining in me," but such "consciousness of myself is by far not a cognition of myself." Since I cannot intuit my spontaneous cognizing activity, "I cannot [cognitively] determine my existence as that of a self-active being." What I *can* cognitively determine is only my existence as a passive empirical phenomenon (*cf.* [3]): TCD of myself yields "not the consciousness of the determining self, but only that of the determinable self" (B407). Thus for Kant, the self that can become an object of empirical cognition(N) is not the self qua spontaneous (determining) cognitive agent.[25] Kant sees a vicious circularity in the attempt to cognize(N) as phenomenal states or processes the a priori conditions for all our cognition(N) of phenomenal states or processes—conditions such as the unity of apperception and the spontaneous acts of thought that unite empirical intuitions in one consciousness (A402; B422).[26]

In light of this, it would be desirable to have a reading that satisfies two desiderata that seem to pull in different directions. On the one hand, we should respect Kant's idea that in acts of TCD we exercise a nonempirical capacity for spontaneously effecting our actual cognitive states. On the other hand, we should respect Kant's point in the B-edition that our *only* nonempirical capacity for causally determining our actuality is the spontaneity of will that is involved in PCD. I suggest that we can satisfy both desiderata if we ascribe to Kant the following view: while the spontaneous activity involved in acts of TCD gives us cognitive mental states as "inner determinations," it lacks a genuine or "true" causality that we exhibit only in acts of PCD.

Kant expresses this view throughout the Dialectic. He says that "human reason shows true causality [*wahrhafte Kausalität*] … where ideas become efficient causes (of actions and their objects), namely in morality" (A317/B374). Through its practical ideas, "pure reason even has causality, to actually produce what its concept contains" (A328/B386). Immediately after claiming that in pure apperception we cognize(W) our capacity for spontaneous "actions and inner determinations," Kant adds that we conceive in ourselves a rational "causality" (only) once we consider "the imperatives that we propose as rules to our powers of execution in everything practical" (A547/B575). In its practical imperatives, reason presupposes that it "could have causality" in relation to our actions, "for without that, it would not expect its ideas to have effects in experience" (A548/B576). The notion that pure practical "reason has causality in regard to appearances" implies a "causality of reason in the intelligible character" (A551/B579)." By

[24]Kitcher's view seems to be endorsed by Grüne (2009, 246–50).

[25]See Ameriks (2000, 289).

[26]Kitcher's response to these passages is that Kant is confused here. In his considered view, the subject of cognitive synthesis is the empirical self that cognizes its own cognitive activity (1984, 122–26). Her argument seems to be as follows: it is a known fact that all representations must be united in one I that thinks. Since this is a *known* fact, "the doctrine of apperception must present phenomenal aspect of the self" (1984, 123). This argument might work if Kant were claiming that all representations *must, as a matter of known empirical necessity*, be united in one empirical consciousness. But, arguably, Kant's doctrine of apperception elucidates an a priori epistemic condition, thus a normative rather than an empirical-cause sense of "must." It is a condition that all representations must meet in order to qualify as objective empirical knowledge, but not a condition that all representations of the empirical self automatically meet (*cf.* Kant's frequent appeal to a *merely* empirical, subjective unity of consciousness, e.g., at B139–40; see also Wuerth [2014, 27]).

"appearances," Kant here means *outer, spatial* appearances: namely, the external effects of moral agency. When I choose to act on the basis of moral imperatives, I expect that I can cause (and witness in experience) physical effects such as repaying my loans, helping others in need, or arriving on campus at the time I promised. Since Kant designates this capacity for PCD as a "true causality," we can infer that according to Kant the capacity for TCD lacks true causality because it is not an efficient cause of what happens in the (empirically) real outside world apart from our mere inner representations of that world.[27]

This distinction between the spontaneity of thought in TCD and the spontaneity of will in PCD is reflected in Kant's definitions of the relevant faculties. He defines the understanding or faculty of cognition as "the capacity to produce representations" (A51/B75), whereas he defines the will or faculty of desire as the "capacity to … produce objects corresponding to representations" (vol. 5, 15) or as "the causality of the power of representation in respect of the actuality of objects" (vol. 28, 1275; *cf.* Bix–x). For example, TCD gives me cognition(N) of an existent table, but via PCD I can bring a table into existence. Kant relies on this contrast on central occasions. For instance, he argues that a deduction of the moral law is impossible because the moral law is conceived as an a priori causal ground of the existence of objects (and such grounds cannot be demonstrated), whereas a deduction of the principles of the pure theoretical understanding is possible because these principles are not causal grounds of objects but, rather, a priori principles for cognizing existent objects that are given to reason from elsewhere (vol. 5, 46–47; see also vol. 10, 131).

Kant's view that only PCD is a "true causality" seems congenial to the ordinary notion that an agent has genuine causal efficacy only if she can produce changes in the outside world and thereby realize her purposes (*cf.* vol. 4, 394). But Kant's view is based on two further considerations.

First, it is significant that Kant defines TCD as a capacity to produce representations *as opposed to* "objects corresponding to representations." Although every representation can be called an object insofar as our consciousness is directed towards it (A189/B234), Kant typically reserves the notion of an object (when he is concerned with sensible phenomena) for the public items of our shared cognition(N) or experience *as opposed to* representations that are private to the mind of individual subjects (A104–5; A191/B236). As Peter Strawson (1966, 88, 24) argues, "objects, in the weighty sense" are "essentially spatial." Part of the reason why a weighty notion of (phenomenal) objects applies only to spatial things is that such things have a permanent existence that our ever-fleeting representations lack (see e.g., Bxli). To be sure, we can also think of nonphenomenal things that are objects in a weighty sense without being spatial. But Kant relies on the contrast between mere representations and genuine objects here as well: when he considers the idea of God that purports to represent a divine being that is "outside" us in a nonspatial (transcendental) sense, he stresses that this idea is a "mere representation" as opposed to a corresponding object (A583/B611), "a mere self-creation of … thinking" as opposed to an "actual entity" (A584/B612). What theology discusses does "not signify the objective relation of an actual object to other things, but only that of an idea to concepts" (A579/B607). Since "mere

[27]One might object that outer objects, as spatial *appearances*, are themselves mere inner representations. This objection rests on a literal reading of Kant's claim that outer appearances are representations, which leads to a phenomenalist version of Kant's idealism (see Van Cleve [1999]). This is (arguably) problematic for various reasons (see Allais [2015] and Allison [2004]). Kant distinguishes between a transcendental and an empirical sense of expressions like "representation," "appearance," "in us," or "outside us" (see e.g., A30/B45; A372–73). Only the empirical sense of "representation" or "in us" entails that something exists literally as a mental state of some particular thinking subject. Physical, spatial things are "representations" or "in us" in a different transcendental sense which designates (roughly) that spatial properties do not exist apart from our way of intuiting things or depend on our representational faculties (see e.g., B69). Kant is committed to denying that spatial objects are (literally) mental states because he draws a contrast between fleeting representations that exist *within* me and permanent spatial objects that exist *outside* myself. (See e.g., Bxxix–xli; when he draws this contrast, he uses the empirical sense of "in/outside myself.")

representations" lack the status of actual (phenomenal or transcendent) objects in a weighty sense, a capacity to produce mere representations lacks causality in a metaphysically weighty sense ("true causality").[28]

Second, for Kant the fact that our theoretical intellect is *only* a capacity to produce representations carries additional metaphysical weight because it marks a decisive contrast between our mind and an infinite, divine understanding. Our theoretical intellect can produce only cognitive representations that depend on the prior existence of their objects. By contrast, an infinite intellect can produce the objects of its representations just by representing these objects (B72; B138–39). Consequently, for a divine intellect there is no distinction between its capacity for theoretical cognition and its will (vol. 5, 403–4; vol. 28, 803, 1054–55, 1272). For a finite mind, a practical faculty of desire that is irreducibly distinct from its theoretical faculty of cognition is the *only* capacity to produce objects in a weighty sense as opposed to mere representations (vol. 10, 130).[29]

These considerations should go some way towards validating my earlier proposal that the distinction between TCD and PCD is a distinction between a cognitive and a causal capacity for self-determination. In Kant's view, our capacity for theoretical cognition is not a "true" causal power because a subject has genuine causality only if she has the ability to produce genuine objects —namely, the external objects of her mere representations. This also confirms that there is no conflict between (4) and (5). When Kant claims (in [4]) that "our actuality is determinable" *only* through PCD, he seeks to emphasize that only our will allows us to produce actual objects of our (purposive) representations whereas the "sheer spontaneity" of thought (in TCD) can give us merely (cognitive) representations of actual objects. Moreover, whereas the spontaneity of will in PCD legislates and executes its purposes independently of empirical desires that result from our affection by external objects, the spontaneity of thought in TCD can achieve its cognitive aims only via empirical intuitions that we receive from existent objects.

Does the fact that our faculty of cognition lacks "true causality" entail that this faculty exhibits only a "relative" type of spontaneity (*spontaneitas secundum quid*), the "freedom of a turnspit" (vol. 5, 96; *cf.* vol. 27, 505; vol. 28, 268) whose actions are ultimately determined by some "foreign" cause? An affirmative answer to this question requires the claim that the causal capacity to produce the objects of one's representations is a necessary condition of absolute spontaneity or freedom. However, Kant defines transcendental freedom in more general terms as a spontaneity to act independently of foreign causes (A533/B561).[30] Our theoretical intellect satisfies this definition if it can perform spontaneous cognitive actions even if these actions lack external effects. Kant affirms that our theoretical intellect does satisfy this definition when he says that we are "completely free" in our cognitive acts of judgment (vol. 4, 290) or that "the understanding … is free and pure self-activity which is determined through nothing other than itself" (vol. 18, 182–83). These quotes date

[28]Kant ascribes even "subjective causality" (causality with respect to mere representations) not to our faculty of cognition but instead to our faculty *of feeling*, which strives to produce or preserve the mental state of pleasure (vol. 28, 741, 815; vol. 20, 206; vol. 5, 220) so that this state has a more than fleeting existence. By contrast, the capacity for TCD does not aim at the production or preservation of mental states: our cognitive activity is intentionally directed at the objects of mental states, but as objects to-be-understood rather than as objects to-be-produced.

[29]The fact that for Kant PCD and TCD involve different types of spontaneous determination raises problems for interpretations that conceive the spontaneity of understanding on the model of the spontaneity of will (see Heimsoeth [1956] and Martin [1969, 207, 212]), and also for interpretations on which our theoretical categories "in a sense cause their objects" so that our theoretical intellect is "a secularized version" of the divine intellect (Forster [2008, 42]).

[30]Allison (1990, 63) and (Kitcher 2011, 246) deny that the spontaneity of thought involves genuine freedom on the basis of A547/B575. In this passage, Kant says that we ascribe to ourselves a free rational *causality* only due to our awareness of practical imperatives. On my reading, Kant here only seeks to stress that our theoretical intellect lacks true causality. This does not entail that our theoretical intellect also lacks transcendental freedom in the more general sense as a spontaneity to act independently of foreign causes. When Kant considers freedom as an ability to initiate a chain of outer events, he is already concerned with a specifically *practical* sense of transcendental freedom that involves the will (A534/B562).

back to the early 1780s, but in in the 1787 B-edition *Critique* Kant still designates our cognitive spontaneity as the "self-activity" of an "intelligence" (B157–59; B278). The notion of self-activity entails absolute freedom (see e.g., A418/B446 and vol. 28, 266–69). Moreover, in the introductions to the 1790 *Critique of Judgment*, Kant ascribes "autonomy" both to practical reason and to the understanding (vol. 20, 225, 241) since the spontaneous legislation of understanding is the source of cognitive laws (vol. 5, 186).

The distinction between a spontaneity that has true causality and a spontaneity that lacks true causality is orthogonal to the distinction between an absolute and a relative spontaneity. The faculty of desire that belongs to a squirrel has "true causality" because it produces external effects (such as digging up a nut), but this causality involves only a relative spontaneity because the inner states (e.g., instincts) that proximately determine the squirrel's behavior are outside the squirrel's control—they are determined by foreign causes such as the evolutionary processes which fix the squirrel's genetic make up. Our faculty of cognition lacks true causality, but it *has* autonomy and thus it can determine its cognitive activity in accordance with cognitive laws (such as the general causal principle) that arise from its own spontaneous self-activity rather than from some external source.[31] These autonomous laws include not only the constitutive laws of the understanding but also the regulative laws that prescribe how "we ought to study nature" and that arise from the "legislation" (A700/B728) of our empirically unconditioned theoretical reason. These laws give our cognitive activity specific aims (such as the systematic unification of all cognitions[N]) and thereby direct us to pursue "the highest purpose of the speculative use of reason" (vol. 4, 350).

Some commentators argue that the spontaneity of cognition is merely relative because our capacity for TCD depends on empirical data that we receive from without.[32] But here we must distinguish between the weaker idea that our capacity to judge is *affected* by empirical conditions and the stronger idea that this capacity is *necessitated or determined* by empirical conditions (A534/B562; vol. 6, 213). The stronger idea would indeed imply a merely relative form of spontaneity. But the notion that acts of TCD are determined by sense perceptions is precisely what Kant denies when he stresses that the input of sensibility leaves us "completely free" to determine for ourselves what our cognitive response to a given perception ought to be (vol. 4, 290). For instance, we must freely make up our mind to consider whether a given perception (that "affects" our mind) warrants a judgment that *p* or rather calls for suspension of judgment.[33] Since our cognitive responses to given data are governed by our representation of cognitive laws and purposes that arise spontaneously from our own autonomous intellect, these responses are not imposed on us by some foreign cause.[34]

[31]Allison (1990, 62) invokes *Reflexion* 5442 (vol. 18, 183) to show that Kant came around to the view that the understanding lacks absolute spontaneity. For Allison, this note even constitutes "the key to [Kant's] change of mind." Kant indeed seems to distinguish here between "logical freedom in actions of reason" and transcendental freedom. But he may wish to deny *only* that "logical freedom" has "true causality." Moreover, it is doubtful whether *Refl.* 5442 expresses Kant's considered view: Kant here defines transcendental freedom as the "complete contingency of actions," which conflicts with published texts (such as vol. 6, 50) where he stresses that absolute spontaneity need not involve any sort of contingency.

[32]See Hanna (2009, 109), Keller (1998, 158–59), Sellars (1970).

[33]One might worry that this implies a kind of doxastic voluntarism which Kant rejects (Chignell 2007; Kitcher 2011, 169). However, a kind of self-determination that is governed by epistemic norms and purposes need not stand under the immediate control of the will or practical reason, as I argue at length in Kohl (2015). See also Buroker (2017).

[34]On some readings, the understanding has absolute spontaneity in its purely *logical* thinking but has only relative spontaneity when it operates under conditions of sensible affection (Keller 1998, 158–59; Sgarbi 2012, 48–49). This implies, oddly, that the understanding shows a more elevated form of spontaneity when it *fails* to realize its cognitive telos. Moreover, the mere fact that a capacity is sensibly affected cannot entail that this capacity has only relative spontaneity: even our absolutely free faculty of choice is essentially affected by empirical desires that yield the "matter" for our free choices (and that make us apprehend the moral law as an *ought*) (4, 412–13; 6, 213, 221).

4.

I have argued that the critical Kant sticks to his claim we can cognize(W) ourselves as noumenal thinking agents who are capable of absolutely spontaneous self-determination. In this final section, I want to expound a tentative suggestion concerning the doxastic status of this claim.

For Kant, every instance of taking some proposition to be true can be classified either as opinion, faith, or knowledge.[35] Opinions are impermissible regarding intelligible properties or beings (A772/B800); this is because opinions for Kant are inherently probabilistic and thus empirical. Hence, our doxastic attitude towards the proposition that our thinking self is a spontaneous intelligence must be either faith or knowledge. My tentative suggestion is that Kant allows that we can know that our noumenal self is a self-active intelligence. My argument here is roughly as follows. Assume that I have cognitive achievements that amount to knowledge. Now assume that I can know that these cognitive achievements presuppose the spontaneous thinking activity of a nonsensible intelligence. If so, I can knowingly infer that "I exist as an intelligence that is … conscious of its faculty for combination" (B158). For Kant, our actual cognition(N) and knowledge of the phenomenal world does presuppose a spontaneous intellectual capacity for the combination of given data, which cannot be ascribed to the phenomenal ("determinable" rather than "determining") self (*cf.* section 3). Hence, reflection on the conditions of our empirical cognition gives us knowledge that our noumenal self is a spontaneous intelligence.

That Kant holds such a view can be confirmed by considering his conclusion to the B-deduction —the categories and the cognitive laws that they yield are "self-thought a priori first principles" (B167) that originate in the pure, spontaneous synthesis of our understanding (A78/B104). The categories are "self-thought" only insofar as they arise from intellectual "self-activity" that is not determined by a foreign cause. Accordingly, at B167 Kant rules out two scenarios where the categories are the product of foreign causes. First, as a priori necessary conditions of all possible experience the categories cannot have "an empirical origin" in natural processes that may be given in experience (and studied in empirical psychology; *cf.* B152). Second, the categories cannot be "implanted in us" by a *super*natural author either. The supposition that the categories might be implanted in us by a supernatural foreign cause that determines us to make categorial judgments is "precisely what the skeptic wishes most" (B168). Once we concede this supposition, we cannot avoid skepticism about the categories because we lack certain criteria for distinguishing between a scenario where the categories originate in a reliable, truthful foreign cause and a scenario where the categories originate in a nonreliable or deceptive foreign cause (vol. 4, 319).[36] The supposition that our judgments which apply the category of causality are determined from without by some designer implies that our causal thought "merely depends on the way in which [our] subject is organized" by something beyond our intellectual self-control, in which case our causal judgments would have merely "subjective necessity" rather than objective validity (B168). Conversely, if we can be certain that our causal judgments have objective validity, then we must be certain that these judgments do *not* merely depend on the way in which our subject is organized by some foreign cause about whose (deceptive or nondeceptive) character we cannot know anything, but rather derive from our own autonomous thinking self-activity. This self-activity reveals itself to us as the source of a priori cognitive laws such as the general causal principle that grounds all our particular causal judgments. These a priori laws cannot qualify as certain knowledge (but instead invite insuperable skepticism) unless we assume that they are "self-thought" by our spontaneous, autonomous intellect. Now these laws do qualify as certain knowledge.[37] Hence, the claim that these a priori laws are self-thought by

[35]For helpful discussion, see Chignell (2007).

[36]Kant here addresses skeptical doubts that concern only the objective validity and truthful origin of a priori, non-empirical concepts. His argument presupposes that we have empirical cognition of objects (*cf.* A93–94/B126–27). Here I follow Ameriks (2003) in reading the Deduction as a "regressive argument."

[37]These laws are apodictically certain (A160–61/B199–200), and apodictic certainty is a mark of knowledge (vol. 9, 66).

our spontaneous, autonomous intellect (*cf.* vol. 5, 186; vol. 20, 225, 241) must also qualify as certain knowledge.[38]

This line of reasoning contains some suppressed premises that I cannot consider here.[39] But Kant's commitment to this line of reasoning can be witnessed elsewhere. For instance, in the second *Critique*, he argues that if we attend to the a priori necessity of first theoretical principles, we can become conscious of our pure, empirically unconditioned understanding as the source or origin of these principles (vol. 5, 30). Our pure, nonempirical self-consciousness pertains to ourselves as things in themselves (vol. 5, 6). Hence, the known a priori necessity of theoretical principles allows us to infer that these principles must originate in our empirically unconditioned noumenal intellect.

In these two passages (B167–68; vol. 5, 30), Kant argues that we have an inferential basis for our knowledge that our theoretical intellect is absolutely spontaneous or self-active. This knowledge is based upon our awareness of representations whose content (an objective a priori necessity) we cannot attribute to an empirically (or otherwise) conditioned intellectual capacity. We might put this by saying that, for example, the a priori causal principle is the *ratio cognoscendi* of our noumenal cognitive spontaneity. (This is precisely what Kant intends to convey when he stresses the parallel between our pure theoretical self-awareness and his claim that the moral law is the *ratio cognoscendi* of our pure, free will [vol. 5, 30].) I suggest that this might exemplify Kant's model of "transcendental reflection," where one considers which of our different sources of cognition is the origin of some (perhaps self-)given representation (A260/B316).[40]

One might object that we cannot know anything about our noumenal self since our noumenal self is not given to us in any intuition and hence cannot be an object of cognition(N).[41] However,

[38]This argument is related to Kant's criticism of the fatalist Schulz (at vol. 8, 114). Many commentators think that Kant's criticism rests on the dubious general premise that, for *any* judgment, one cannot rationally endorse it if one is causally determined to accept it (Wood [1999, 177–80], Rosefeldt [2000, 169–84]). However, in my view Kant argues more specifically (and forcefully): we cannot rationally endorse an *a priori* principle that expresses an objective *necessity* (like the general causal principle, which Schulz accepts) while assuming that our judgments are (as our volitions, on Schulz's view) empirically determined. The latter assumption implies that our judgments and their propositional contents are contingent on the operation of "merely subjectively determining causes that can change subsequently" (vol. 8, 14), i.e., on the operation of some mutable, rationally arbitrary psychological mechanism (like Hume's blind habits). In this scenario, our judgments express only a "subjective necessity" (B168) and cannot lay claim to an objective, timeless a priori necessity (see e.g., vol. 18, 176; vol. 20, 238). For Kant "the understanding … is free and pure self-activity" *because* "without this … *immutable* spontaneity, we would cognize nothing *a priori*" (vol. 18, 182–83; emphasis mine).

[39]For instance, the argument requires the antiexternalist premise that even if there were a supremely reliable causal link between our concepts and their objects (e.g., if a benevolent God were the common cause of our concepts and their objects), this would *not* give our concepts objective validity. Stang (2016, 192–94) argues that Kant's objection to divine preformation theory cannot rule out a reliabilist view on which we might possess knowledge that is grounded in divine imposition. Rather, Stang suggests, Kant sets aside the divine imposition theory he associates with Crusius because it is circular to rely on claims about a divine being in the context of addressing the question of whether our categories can provide a priori cognition of noumena. However, Kant does not address this question at B167–68 (he has already ruled out categorial cognition of noumena at B147–49). Rather, he argues that the categories must be originally acquired via spontaneous self-activity in order for them to enable objective cognition *of phenomena*. Kant's complaint that on the preformation theory our categorial judgments rest on a "subjective necessity" shows that he classifies this theory alongside Hume's view, which Kant routinely calls as a "subjective necessity" theory as well (B5; A94/B127; vol. 5, 12–13). What both theories have in common is the upshot that we think in categorial terms *just because* our mind is organized by some external principle that, as such, is completely removed from our conscious intellectual self-control. For Kant, this upshot is incompatible with the very idea of a priori knowledge, which must be grounded in our own reason (vol. 9, 70) and which requires a consciousness of objective necessity and certainty (vol. 9, 71).

[40]Against my inferential reading, one might object that in Kant's view we have an *immediate*, noninferential awareness of our activity as spontaneous noumena. For a defense of this view, see Wuerth (2014). Wuerth's view rests on the claim (which he attributes to Kant primarily on the basis of unpublished texts) that we possess "an intuition of our self" which is "the only intuition of a thing in the noumenal sense" (2014, 77). Wuerth is right to note that (given Kant's link between immediacy and intuition) any immediate (noninferential) awareness of ourselves must be intuitive (rather than purely conceptual). However, an intuition of our noumenal self would have to be a nonsensible, intellectual self-intuition, and Kant's critical view denies that we can have such intuition (see e.g., B159).

[41]See Wuerth (2014, 173).

Kant sometimes allows for knowledge of *x* that is not grounded in cognition(N) of *x*.[42] Consider the belief that there are things in themselves. Arguably, this belief amounts to knowledge: we know (through sense perception) that certain objects exist, and we know (through philosophical reflection) that we experience only the mind-dependent sensible constitution of these objects. Hence, we can knowingly infer that these objects also have some nonsensible constitution.[43] Here the knowledge that there are nonsensible aspects of reality is justified as a certain entailment of our knowledge regarding the sensible aspects of reality: "… beings of the understanding certainly correspond to sensible beings …" (B308); "… beyond the sensible world there must necessarily be found something that is thought only by the pure understanding …" (vol. 4, 361). Since we lack cognition(N) of things in themselves, we cannot turn the abstract knowledge that there *are* nonsensible aspects of reality into determinate knowledge of what these nonsensible aspects are like or how things are constituted in themselves, i.e., in their intelligible character (B149): "… we … cannot know anything determinate about … intelligible beings" (vol. 4, 315).

Similarly, even though we can have abstract knowledge that our spontaneous noumenal self is responsible for our cognition(N) and for our determinate knowledge of the sensible world, the fact that we cannot have cognition(N) of our spontaneous thinking self prevents us from acquiring determinate knowledge of how our spontaneous thinking self is constituted. This is the lesson Kant wishes to draw from the Paralogisms: we cannot cognize(N) or know in any determinate fashion the ontological structure of the intelligible subject of "sheer spontaneity," e.g., whether it is a simple (and thus incorruptible) or a complex being.[44] Accordingly, Kant insists (at B158) that the ascription of spontaneity to the thinking self in apperceptive self-cognition(W) does not suffice to (cognitively) "determine its existence" as a noumenon.

One might worry that if Kant did grant that we know on a theoretical basis that our faculty of cognition has absolute spontaneity, then it would follow that we also know on a theoretical basis that our moral will has absolute spontaneity. This would be fatal implication given Kant's insistence that we cannot theoretically demonstrate the (real) possibility of our transcendental freedom of will (A557–58/B585–86). Robert Pippin (1987, 473) suggests that the ascription of absolute spontaneity in one (epistemic) context does not entail that one must also ascribe such spontaneity in other (moral) contexts. For Patricia Kitcher (1990, 253), this response is too weak since it provides no reason that blocks the generalization from one context to the other. My interpretation strengthens Pippin's point because it supplies a twofold, principled reason why knowledge of absolute cognitive freedom does not entail knowledge of absolute moral freedom.[45] First, for Kant the spontaneity of thought that is involved in TCD does not include *the* most controversial, metaphysically weighty feature that we attribute to a spontaneous moral will—namely, a nonsensible form of efficient causality that allows us to produce actual external objects which are distinct from mere representations. Second, Kant argues that we must posit the autonomous spontaneity of our theoretical intellect as an object of indeterminate cognition(W) and knowledge because the exercise of such spontaneity is— *unlike* the exercise of moral spontaneity—presupposed by our actual cognition(N) and determinate knowledge of the phenomenal world. For these reasons, we know that we possess

[42]For this view see Chignell (2014), Schafer (forthcoming), and Watkins and Willaschek 2017.

[43]See Allais (2010, 15).

[44]Thus, if Kant permits us knowledge that our cognizing self has noumenal spontaneity, he is not thereby committed to granting the determinate self-knowledge to which rationalist theories aspire. See Ameriks (2000, 287–90).

[45]Rosefeldt (2000, 167–68) suggests that we cannot theoretically cognize our freedom of will, as opposed to our freedom of thought, because we cannot cognize whether our actions are done from duty (based on the pure moral law). However, this seems to imply the Sidgwickian view that our actions are free only if they are done from duty, an implication that Rosefeldt endorses (2000, 159, 165). This view is philosophically implausible and textually indefensible. In both the first Critique (including the 1781 A-edition) and the second *Critique*, Kant's chief examples of free actions are immoral actions such as lies and thefts (A554/B582; vol. 5, 95). See Wuerth (2014, 236–54) for critical discussion of the Sidgwickian view.

absolute cognitive spontaneity, but we cannot know (at least not without further practical-moral arguments) that we also possess the truly causal absolute spontaneity of a noumenal moral will.

Acknowledgments. For very helpful comments on this essay, I am grateful to the editors and two anonymous referees of this journal.

Markus Kohl is assistant professor of philosophy at the University of North Carolina, Chapel Hill. He obtained his PhD in philosophy in 2012 from the University of California, Berkeley. From 2012 to 2017, he was assistant professor of philosophy at the University of Tennessee, Knoxville.

References

Allais, Lucy. 2010. "Transcendental Idealism and Metaphysics: Kant's Commitment to Things as They Are in Themselves." In *Kant Yearbook 2010: Metaphysics*, edited by D. H. Heidemann, 1–33. Berlin: De Gruyter.

Allais, Lucy. 2015. *Manifest Reality. Kant's Idealism and His Realism*. Oxford: Oxford University Press.

Allison, Henry. 1990. *Kant's Theory of Freedom*. Cambridge: Cambridge University Press.

Allison, Henry. 1996. *Idealism and Freedom*. Cambridge: Cambridge University Press.

Allison, Henry. 2004. *Kant's Transcendental Idealism*. New Haven, CT: Yale University Press.

Ameriks, Karl. 2000. *Kant's Theory of Mind*. Oxford: Oxford University Press.

Ameriks, Karl. 2003. *Interpreting Kant's Critiques*. Oxford: Oxford University Press.

Buroker, Jill Vance. 2017. "Kant on Judging and the Will." In *Kant and the Philosophy of Mind*, edited by A. Gomes and A. Stephenson, 189–203. Oxford: Oxford University Press.

Chignell, Andrew. 2007. "Belief in Kant." *Philosophical Review* 116: 323–60.

Chignell, Andrew. 2014. "Modal Motivations for Noumenal Ignorance: Knowledge, Cognition and Coherence." *Kant-Studien* 105 (4): 573–97.

Chignell, Andrew. 2017. "Can't Kant Cognize Himself? Or, a Problem for (Almost) Every Interpretation of the Refutation of Idealism." In *Kant and the Philosophy of Mind*, edited by A. Gomes and A. Stephenson, 138–57. Oxford: Oxford University Press.

Forster, Michael. 2008. *Kant and Skepticism*. Princeton, NJ: Princeton University Press.

Grier, Michelle. 2001. *Kant's Doctrine of Transcendental Illusion*. Cambridge: Cambridge University Press.

Grüne, Stefanie. 2009. *Blinde Anschauung*. Frankfurt am Main: Klostermann.

Hanna, Robert. 2009. "Freedom, Teleology, and Rational Causation." In *Kant Yearbook 1/2009,* edited by D. H. Heidemann. Berlin: De Gruyter.

Heimsoeth, Heinz. 1956. *Studien zur Philosophie Immanuel Kants*. Köln: Kölner Universitäts-Verlag.

Hogan, Desmond. 2009. "Three Kinds of Rationalism and the Non-Spatiality of Things in Themselves." *Journal of the History of Philosophy* 47: 355–58.

Keller, Pierre. 1998. *Kant and the Demands of Self-Consciousness*. Cambridge: Cambridge University Press.

Kitcher, Patricia. 1984. "Kant's Real Self." In *Self and Nature in Kant's Philosophy*, edited by Allen Wood, 113–47. Ithaca, NY: Cornell University Press.

Kitcher, Patricia. 1990. *Kant's Transcendental Psychology*. New York: Oxford University Press.

Kitcher, Patricia. 2011. *Kant's Thinker*. New York: Oxford University Press.

Kohl, Markus. 2015. "Kant on Freedom of Empirical Thought." *Journal of the History of Philosophy* 53 (2): 301–26.

Kohl, Markus. 2016. "Kant on Freedom, Idealism, and Standpoints." *Archiv für Geschichte der Philosophie* 98 (1): 21–54.

Longuenesse, Beatrice. 1998. *Kant and the Capacity to Judge*. Princeton, NJ: Princeton University Press.

Longuenesse, Beatrice. 2017. *I, Me, Mine*. Oxford: Oxford University Press.

Marshall, Colin. 2010. "Kant's Metaphysics of the Self." *Philosophers Imprint* 20: 1–21.

Marshall, Colin. 2013. "Kant's One Self and the Appearance/Thing-in-itself Distinction." *Kant-Studien* 104 (4): 421–41.

Martin, Gottfried. 1969. *Immanuel Kant*. Berlin: De Gruyter.

Pippin, Robert. 1987. "Kant on the Spontaneity of Mind." *Canadian Journal of Philosophy* 17: 449–75.

Puls, Heiko. 2015. "*Kant über die äusserste Grenze aller praktischen Philosophie – Ein Kommentar zur Sektion 5 der Grundlegung.*" in *Kants Begründung von Freiheit und Moral in Grundlegung III*, edited by Dieter Schönecker, 157–214. Münster, Ger.: Mentis.

Rosefeldt, Tobias. 2000. "*Das logische Ich: Kant über den Gehalt des Begriffs von sich selbst.*" Hamburg: Philo.

Schafer, Karl. Forthcoming. "Kant's Conception of Cognition and Our Knowledge of Things-in-Themselves." In *The Sensible and Intelligible Worlds: New Essays on Kant's Metaphysics and Epistemology*, edited by K. Schafer and N. Stang. New York: Oxford University Press.

Schönecker, Dieter. 1999. *Kant: Grundlegung III. Die Deduktion des kategorischen Imperativs*. Freiburg, Ger.: Alber.

Sellars, Wilfrid. 1970. "… This I or He or It. (The Thing) Which Thinks …" *Proceedings of the American Philosophical Association* 44: 5–31.
Sgarbi, Marco. 2012. *Kant on Spontaneity*. London: Continuum Publishing.
Stang, Nicholas. 2016. *Kant and the Metaphysics of Modality*. Oxford: Oxford University Press.
Strawson, Peter. 1966. *The Bounds of Sense*. London: Methuen.
Van Cleve, James. 1999. *Problems from Kant*. New York: Oxford University Press.
Watkins, Eric, and Marcus Willaschek. 2017. "Kant's Account of Cognition." *Journal of the History of Philosophy* 55 (1): 83–112.
Wood, Allen. 1999. *Kant's Ethical Thought*. Cambridge: Cambridge University Press.
Wuerth, Julian. 2014. *Kant on Mind, Action, and Ethics*. Oxford: Oxford University Press.

Cite this article: Kohl, M. 2020. Kant on Cognizing Oneself as a Spontaneous Cognizer. *Canadian Journal of Philosophy* 50: 395–412, doi:10.1017/can.2019.44